Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an

advanced degree from Emory University, I hereby grant to Emory University and its agents the

non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole

or in part in all forms of media, now or hereafter known, including display on the world wide

web. I understand that I may select some access restrictions as part of the online submission of

this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or

dissertation. I also retain the right to use in future works (such as articles or books) all or part

of this thesis or dissertation.

Signature:

_____   _____

                    Tianchi Zhang   Date

Apply Group ICA Method into fMRI Analysis

Using GIFT Package and Regression Models

By

Tianchi Zhang

MSPH

Emory University

Rollins School of Public Health

Department of Biostatistics

_____ [Thesis Advisor's signature]

Ying Guo

_____ [Reader's signature]

Jian Kang

Apply Group ICA Method into fMRI Analysis

Using GIFT Package and Regression Models

By

Tianchi Zhang

B.S., Nankai University, 2012

Thesis Committee Chair: Ying Guo, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Department of Biostatistics

2014

Apply Group ICA Method into fMRI Analysis

Using GIFT Package and Regression Models

By: Tianchi Zhang

Abstract

This paper focused on the application of group independent component analysis (ICA) method into fMRI analysis and evaluating covariate effects on independent components in group ICA. We first applied GIFT package developed by medical image analysis laboratory (MIALab) for group ICA analysis on our dataset. To evaluate the covariate effects, we considered the Mancovan package in GIFT. Our major issue with Mancovan is that it works on group IC maps instead of subject-specific IC maps for evaluating the covariate effects. To address this limitation, we propose a multivariate regression method based on subject-specific maps. We evaluated the performance of our regression approach using simulation studies that generated multi-subject fMRI data with simulated covariates. The accuracy of the method is evaluated by the Mean square error (MSE) based on the difference between the true and estimated covariate parameters. Our results show that the proposed regression method may provide a good way to assess covariate effects on spatial distributions of ICs extracted from fMRI data.

Apply Group ICA Method into fMRI Analysis

Using GIFT Package and Regression Models

By

Tianchi Zhang

B.S., Nankai University, 2012

Thesis Committee Chair: Ying Guo, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Department of Biostatistics

2014

Acknowledgements

I want to thank the faculty, advisors, and staff of the Biostatistics Department at Rollins School of Public Health for the dynamic two years of learning that I have had. This thesis is only a sample of the vast knowledge that was attained and applied through my study at Rollins. I would especially like to thank Professor Ying Guo for all of her advice and support to help me write this thesis. Also a special thanks to Jian Kang for taking the time to read my thesis.

At last I want to give a special thanks to my parents for their loving support and encouragement that motivated me to pursue a degree at the graduate level.

Table of Contents

<center>**<u>Chapter I</u>**</center>

Introduction

Overview and background

A lot of neuroimaging techniques have been developed to investigate different aspects of brain procedures, one is functional magnetic resonance imaging (fMRI). fMRI is a kind of MRI. MRI is mainly used to generate structural images of organs, among them brain is the most widely used part. It can also be used to reveal information on the different states of tissues, the blood flow measurements for tissues and organs and perfusion (DTI). fMRI aims to measuring changes via changes of blood amount, or changes in the concentration of oxygen. The last technique is known as the BOLD contrast. It is the most popular technique in human functional neuroimaging.

The statistical analysis for neuroimaging data is difficult because it's very complex. One reason is the structure of its spatial and temporal correlation. Another reason is its high-dimensionality, this must need a lot of time to compute and brings heavy burden to the computation. There are many methods now to make the computation faster and easier, the main idea for these method is reducing the dimension. And a very popular method now is Independent component analysis (ICA).

Organization of Human Brain

The human brain is a complex organ and it contains over 100 billion nerve cells, we also call them neurons. Each neuron receives and sends signals to other neurons. There are several

basic components for a neuron, they are cell body, dendrites and axons. The cell body controls the entire neuron activity, the dendrites receives signals from other nerve cells, and axons sends signals to others. They transfer information to each other through electrical impulses called "action potentials". And this process needs energy to support, the energy is from glucose and oxygenated-hemoglobin. Thus the brain is more activated, the rate of metabolism is higher and oxygenated-hemoglobin concentration is increasing.

The whole brain consists of left hemisphere and right hemisphere. There are four main sections depending on different functions. They are frontal lobe, temporal lobe, parietal lobe and occipital lobe. Frontal lobe mainly takes charge of making decision and solving problem, temporal lobe regulates memory, emotion and language, parietal lobe receives and processes information from the body and occipital lobe handles information related to vision. And there are 48 brain regions in total based on one classification of properties. In another definition, there are 90 regions. The most important thing needed to know about human brain is that different regions in human brain are responsible for different activities.

fMRI

fMRI is a technique that measures brain activity as a function of the hemodynamic response to experimental stimuli since the early 90s. It uses blood oxygenation level dependent (BOLD) contrast based on the difference between magnetic properties of oxygenated (diamagnetic) and deoxygenated (paramagnetic) blood. The procedure goes in this way: When there is an action in the brain, which means the neurons are activated, there will be a change in the blood

flow and oxygenation, then this causes a change in a MR decay parameter T_2^*. T_2^* will be weaker in areas where oxygen concentration is high, while in areas the concentration is low, there will be a weaker T_2^* signal. Using the contrast, fMRI can measure differences in brain activities over time.

fMRI Data

The MRI signal is acquired as a quadrature signal. This means two orthogonal "detectors" are used to get the MRI signal. The outputs from this system are usually in a complex form with one output as the real part and the other as the imaginary part. The time-domain data can be seen as the spatial-frequency representation of the image data. It is common to take the magnitude of this data prior to performing any fMRI analysis. In a fMRI study, one subject is scanned multi times, so fMRI data is represented as a query of 3D scans over time, which is also can be seen as a 4D data (time can be seen as the 4th dimension). The basic unit to make up the whole 3D scan is voxel, can be seen as the 3D extension of pixels. The BOLD signal is measured at each voxel, the value represents for the level of brain activity at that location. If each signal is recorded for a single voxel, a time course of brain activity for a single voxel can be got. And if we use V to represent the number of voxels and T to represent the number of scans during the fMRI study, the data can be seen as a two-dimensional V by T matrix.

Usually fMRI data is very large because of thousands of voxels in each scan plus hundreds of scans during the study. The large and complex dataset make computation more complex. This leads us to apply some statistical methods to simplify the complication.

In our study, the data we used are in NIfTI format. NIfTI format is new analyze-data format, the full name for NIfTI is Neuroimaging Informatics Technology Initiative. It defines coordinates relating voxel index (i,j,k) to spatial location (x,y,z). It is usually stored as a single file (.nii) or as two files combined (.hdr and .img).

Chapter II

Methodology and Experiment

Review of ICA Methods

Independent component analysis (ICA) is a promising data analysis method that is being increasing applied to fMRI data. Independent component analysis is a computational method for separating a multivariate signal into additive subcomponents by assuming that the subcomponents are non-Gaussian signals and that they are all statistically independent from each other. Independent component analysis can be applied to fMRI data to separate either spatially or temporally independent sources. Spatial ICA can find systematically non-overlapping, temporally consistent brain regions without controlling the temporal domain. One issue in fMRI study is that the temporal dynamics of fMRI experiments are difficult to study due to the lack of a good brain-activation model [2]. ICA can revel dynamics for which a temporal model is not available.

Here is a simple explanation for ICA method, which comes from the first paper to introduce ICA method [5]. In a closed room, there are two people talking at the same time, and there are two microphones placed in the room, too. They are used to record signals over time. Use $s_1(t)$ and $s_2(t)$ to represent the two speakers, which can be seen as two sources. And the recorded signals for the two microphones are marked as $x_1(t)$ and $x_2(t)$. Each signal is a weighted sum of the two sources. This can be expressed as:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

The goal in this example is to estimate the original source signals $s_1(t)$, $s_2(t)$ using the recorded signals $x_1(t)$ and $x_2(t)$.

And before the conduction of ICA method, there needs to be preprocessing procedures. This can make the ICA estimation simpler and better conditioned.

The characteristic of BOLD signal is a main reason to apply independent component analysis in fMRI study. BOLD signal is very weak and noisy. To subtract useful signals regardless of noisy signals, ICA is a perfect way. And the goal of ICA is to use the observed fMRI signal to estimate the statistically independent components and their corresponding time courses. Use $X_{T \times V}$ to represent the matrix of observed fMRI data, $S_{q \times V}$ to represent source component matrix. T is the number of time points, V is the number of voxels and q is the number of components. $A_{T \times q}$ can be seen as the "mixing" matrix.

$$X_{T \times V} = A_{T \times q}S_{q \times V}$$

Introduction to Group ICA methods

ICA has been proven to analyze single-subject fMRI data very successfully, and recently extended to multi subjects. This process is more complicated. One reason for the more complicated process is that ICA does not draw good inferences about groups of subjects. In comparison with regression analysis, an example is that when using the regression model, the

researchers specify parameters of interest, then from these parameters they can draw

inferences about the group well because all subjects in the group share the same parameters.

But in ICA, different subjects in the group may have different time courses and different spatial

maps, which can be seen as they have different parameters, and they will be sorted differently,

so it's not that easy to draw inferences about multi-subject data using ICA.

A key problem in group ICA is how to combine different components across subjects. Consider

the simple example which has two subjects, and there are two time points. Use x and y to

represent the two subjects, the data at two time points are $d_x = [x_1 \; x_2]$, $d_y = [y_1 \; y_2]$. We set

a vector to combine the data at two time points together. And then combine the two subjects

into one vector, we have

$$d = [d_x^T \; d_y^T]$$

Then for a simple consideration, we set the number of sources as 2. The "mixing" matrix is:

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{bmatrix} = [A_x \; A_y]$$

Here A_x and A_y represent for the "unmixing" matrix for x and y. And an estimate of the original

source is $\hat{s} = Ad$. And we assume independence exists in $\hat{s}$. The independence consists of

between-subject independence and within-subject independence. Then we can composite $\hat{s}$ as

$$\hat{s} = \begin{bmatrix} s_{11} + s_{12} \\ s_{21} + s_{22} \end{bmatrix}$$

Here we set $s_{11} = \alpha_1 x_1 + \alpha_2 x_2$, $s_{21} = \beta_1 x_1 + \beta_2 x_2$, $s_{12} = \alpha_3 y_1 + \alpha_4 y_2$, $s_{22} = \beta_3 x_1 + \beta_4 x_2$.

Because ICA minimizes the dependence among rows of $\hat{s}$, the dependence of s_{1i} and s_{2i} (i=1,2)

is more affected by the data in each subject. This results the parameters for each subject are determined the subject's observations. So the matrix for each subject is separable, and the data reconstructed is more likely to be a combine of data within subjects. It minimizes the across-subject effect, the key problem isn't a problem now.

Currently there are different multi-subject ICA methods for analyzing fMRI data. These methods have difference in how the data are preprocessing before conducting ICA and the way group and subject estimates are computed. One method is estimating and comparing group temporal responses and common spatial maps [9].

And group ICA methods can be summarized into five categories [8]. One category is using single-subject ICA and then combining the output into a group using clustering approaches. But most methods compute the group data as a whole. In this way both temporal concatenation and spatial concatenation can be examined. And the advantage is that for one component the comparison of subject differences is very straightforward because when one ICA is conducted, it can divided into different parts based on different subjects. And temporal concatenation performs better for fMRI data since the temporal variations in the ICA time courses are much larger than the variation in the spatial maps. So temporal concatenation is now widely used for group ICA method to analyze fMRI data. And it is used in the group ICA of fMRI toolbox (GIFT), which we used in our study.

Description of data

In this study, there are 35 subjects in total. For each subject there is a corresponding NIfTI file recording the fMRI data. It can be seen as a multi-dimensional matrix. For example, the dimension for the first NIfTI file is 47×56×46×216. This means the 3D coordinate for the brain of the first subject is (47, 56, 46). And the number of time points in the fMRI time series is 216. This number retains the same for all subjects.

The covariates in our study were: Age for each subject, PTSD diagnosis result for each subject, result of a preliminary investigation questionnaire called TEI, result of Childhood Trauma Questionnaire (CTQ), result of a 17-item semi-structured interview that assesses the presence and severity of DSM-IV PTSD symptoms (PSS), result of the Beck Depression Inventory questionnaire (BDI) and result of an assessment for ED symptoms (EDS). Among these covariates, all were continuous variables except for PTSD. We recoded this covariate as 1 for Yes and 0 for No.

Experiment Procedures:

In this project we mainly used the Group ICA Toolbox (GIFT) package in Matlab developed by Medical Imaging Analysis Lab (MIALab). This package includes a lot of algorithms for independent component analysis and source separating of group functional magnetic resonance imaging data.

The main steps of conducting ICA analysis using GIFT package includes using PCA for data reduction for every subject, ICA decomposition to subtract components and then back-construction to get subject-specific maps and time courses.

The first issue we needed to conduct by ourselves was estimating the number of independent component and setting this number up. Components were estimated using the MDL (Minimum description length) criteria. The numbers of components were calculated for each subject, then the mean was calculated across all 35 subjects. The result for the mean number of component was 99. It was a very large number. If we used this number to go on with the analysis, it would increase the analyzing complication a lot. Thus we set the number of components to 20, which was a common-used number in this analysis.

The default brain mask didn't work well on our data. Therefore, we took the following approach to build our own binary mask based on the data using the AFNI software. We checked the validity of the developed mask using Micron and Fslview. A binary mask is made with simple two-level values 0 and 1. The main idea is using all non-zero voxels to create a mask.

In ICA analysis step, we got component spatial maps as well as time courses. Group map and subject-specific maps were also collected in this step to get prepared for following analysis. We

could classify the components by sorting them either spatially or temporally. Temporal sorting provides a way to compare the model's time course with the ICA time course. Statistical tests could also be used here to compare the components. Two-sample t-test could be used to compare single component between subjects. And we could also generate regression parameters for performing tests between components or groups or evaluating the relatedness between tasks of the components.

For purpose of seeking the relationship between these ICA analysis results and other covariates recorded for each subject, we needed to perform a multivariate analysis of covariance. This analysis was conducted with using Mancovan package also developed by MIALab. And multivariate tests were supposed to be done first to determine the significant covariates that were later used in the univariate tests.

An important step of Mancova analysis was setting the design matrix. A design matrix was determined by the covariates of interest, interaction terms generated by the covariates as well as nuisance terms in consideration of spatial normalization. Each term comprised a vector, and each vector must have equal length to the number of subjects in the previous ICA analysis. Here the length was 35. To meet this criteria, we replace missing values with 0. Results from independent component analysis were also included as input for Mancovan function.

Spatial maps were produced to show the result of Mancova analysis. Here the spatial maps were called T-maps which were based on t-statistics. Here we set the significant level for p value as 0.05. We also got multivariate results and the significant ones could be plotted in a figure. The more significant the covariate was in the component the darker the color should be.

But we found the multivariate result from Mancova analysis isn't ideal. And this package conducted multivariate analysis based on group maps, not subject-specific maps. So we conducted multivariate analysis on our own. Our thought was using regression methods.

For the proposed methods, we first obtained subject-specific IC maps using the back-construction method in GIFT and we then performed a multivariate regression on the subject-specific maps. The multivariate regression model was in a general form of the formula below.

$$S = \beta X + \varepsilon$$

It's similar to the multiple regression model, but pay attention here the outcome variable S is a matrix instead of a vector in multiple regression model. Each row in S represents for one subject, for each subject we have:

$$S_i = \beta X_i + \varepsilon$$

S_i is a row vector instead of a single number. Each element (column) in the vector represents for one component. This is why we use multivariate regression models, we have multi components. Further explanation will be given in simulation part.

And then we conducted a simulation study to evaluate the performance of the proposed regression approach.

Simulation Study

The simulation study was conducted in this way: First we generated the data for component

and covariates. We used X_i to represent for the covariates for each subject. Here X_i was a

$p \times 1$ matrix, p was seen as the number of covariates. $\beta(v)$ was used to represent for the

regression coefficient matrix, the dimension was $q \times p$. q was the number of independent

components. Then, we generate the spatial source signals for each subject from the following

model. The dimension for source was $q \times 1$. ε was the error term. ε was also a $q \times 1$ matrix

with each term following normal distribution $N(0, \sigma^2)$.

$$S_i = \beta(v)X_i + \varepsilon$$

Then we generated a "mixing" matrix to generate the fMRI results. This was similar to the

formula in ICA method introduction mentioned before. We used y_i to represent for the fMRI

results for subject I and A_i to represent for the "mixing" matrix (see below). To generate the

temporal mixing matrix, we used estimated time series from real fMRI data we had in our

study.

$$y_i = A_i S_i$$

After getting the fMRI result, we applied the ICA method to get $\hat{s}_i$, called "inferred"

component. We wanted to compare the "inferred" component with original component. We

applied the proposed regression method to obtain an estimate for the covariate effects,

marked as $\hat{\beta}(v)$. Specifically, $\hat{\beta}(v)$ was estimated using the ordinary least square estimation

(OLS) method. OLS was estimating the unknown parameter by minimizing the sum of squared

differences between the observed responses in the dataset and the predicted responses. The

estimated $\hat{\beta}(v)$ was calculated by:

$$\hat{\beta}(v) = (X'X)^{-1}X'\hat{S}$$

X and $\hat{S}$ are the covariate matrix and estimated sources concartenated across subjects. After

getting $\hat{\beta}(v)$ we compared it with the original coefficient $\beta(v)$.

We run this simulation for 100 times. And mean square error (MSE) was used to quantify the

difference between $\hat{\beta}(v)$ and $\beta(v)$. MSE was the average of the squares of errors. The error

was calculated for each simulation. Since both $\hat{\beta}(v)$ and $\beta(v)$ were matrixes, we needed to find

a way to calculate the error. The mean was meaning calculating the average error across all

voxels. For each voxel, calculate the difference between $\hat{\beta}(v)$ and $\beta(v)$. The difference was

calculated by taking the square root of the sum of all differences between the elements at

same position of these two matrixes. Then take a square of the difference and sum up, average

by the number of voxels. The formula was shown below.

$$MSE = \frac{1}{V}\sum \left\| \hat{\beta}(v) - \beta(v) \right\|^2$$

We recorded the MSE for each simulation, and calculated the mean and standard deviation for

all times of simulation.

In this simulation study, we set p=2, q=3, the variance for the error term $\sigma^2 = 0.5$. And there

were other parameters we also needed to set to complete the simulation. We assumed there

were N=10 subjects in each group. With N and p we could generate X (10×2) to represent for

the covariate matrix for the whole group. Each element was generated from uniform

distribution but with different parameter settings for different columns. And we assumed the

dimension for each imaging file was 25×25×4, so the number of voxels was the product result

400, marked as V. With p, q and V we could generate the covariates regression coefficient

matrix, marked as β. We also plotted β to get a direct comparison with the estimated

coefficient. After applying ICA and regression method, we could get the estimated coefficient $\hat{\beta}$

and plotted it out.

Chapter III

Results and conclusions:

Using GIFT package we can get the component map (Figure 1- Figure 5) showing all components of mean for all subjects. Since the brain is actually a 3D imaging and it was difficult to be shown on the paper, it was shown by a series of 2D images which was on $X - Y$ platform. The numbers on the bottom left corner of each image represented for the Z value, which represented for the vertical coordinate. The time course for each component was displayed on the top of each figure. And the component was not sorted here.

Table 1 showed which component was most significant with each covariate in terms of spatial maps and time-course spectra based on group maps from ICA analysis. Here we mainly looked at the spatial map part. We found component 13 was significantly related to three covariates, TEI, EDS and PSS. And for the most important covariate PSS, the most significant one was the eighth component.

The T-maps for the first component were shown in Figure 9 and Figure 10 as an example. Each T-map contained three small maps, which were three directions of 2D views. The red and yellow parts in each map showed the significant regions. The number under each small map represented for the coordinate in each direction. By combining different maps together we could exactly know where the significant region located. Figure 8 showed the multivariate results when we only used PTSD, PSS and the interaction term PTSD*PSS to create the design matrix. The yellow parts in the matrix of dimensions covariates by components are supposed to

show the significant results. From this figure we found just PSS (the second row) was significant in the eighth component (the second column). Just one significant result made no sense.

In the simulation study, the original regression coefficient was plotted in Figure 9. The estimated coefficient from the regression was plotted in Figure 10. We noted there were three columns, each column represented for a component. And there were four rows, it was the same number as the last dimension number we set for the voxel. Here the estimated coefficient figure was plotted by the average estimation, because if just one estimation was taken to make a comparison it would be inaccurate. Compare these two figures we found they were very similar except for the estimated coefficient figure appeared to be a little and had more white dots. And the mean for the MSE was 0.33, the standard deviation for the MSE was 0.21. These also showed there was not much difference between the original coefficient and the estimated one. This showed regression was also a good method for multivariate analysis for imaging data, but it needed to be combined with other methods like ICA.

During the experiment procedure we also found some tips and points needed to pay attention to for the GIFT package such as including different subjects in different folders and building a mask instead of using default mask calculated automatically. For the first issue, from our point of view the problem is that if we load NIfTI files and put them all in one folder, the package will see them as from one subject automatically. For the second issue, the default mask is calculated using all the file for subjects or only the first file for each subject depending on the "Default_MASK_OPTION" variable in the defaults. There may be problems in the calculation method when applying to our subjects, but we are not sure where the problem is.

Chapter IV

Discussion:

In the first step of using GIFT package, the mask we created was a simple binary mask based on the first subject. There might be problems with these because a binary mask might not be the best-fitted mask for our data. In building this mask we used binomial distribution, which was a discrete distribution, but the data was seen as continuous. We could try to build a mask using other continuous distributions such as exponential distribution and gamma distribution. We also needed to build masks based on other subjects and then compare the masks to see whether they were very similar.

For the Mancova analysis, we included all covariates and all the interaction terms generated by them, there were too many terms. A selection procedure should be conducted. To ensure all covariates vectors have the same length we replaced missing values with 0. This might be also the problem for the unexpected bad result. Another reason might be we didn't include the nuisance predictors so we didn't ensure the normalization. Further work should be focused on how to creating an appropriate design matrix. A simple idea dealing with missing values deleting them, but there must be many problems. If we just delete the missing values, we also need to delete some values of other covariates for equal length. Which value to delete is also difficult to determine. Another idea is replacing them with the mean of all other valid values for each covariate. We need to test the efficiency for this idea and try other methods to find a good way to replace missing values.

This simulation study used the simplest regression model, though this showed a not bad result, the difference between true and estimated coefficient could be smaller. The white dots could be seen as noises, showing the estimation was not that ideal. An idea for solving this problem is to apply more complicated regression results to reduce the interaction. Another improvement can be considered in replacing OLS method with more complicated method such as weighted least squares (WLS) or generalized least squares (GLS). WLS is a good option here because it can deal with heteroscedasticity, this is very common in real data. GLS is a good method when there is certain correlation between the observations, this is also very common for real data. So WLS and GLS are more robust for estimating. In our simulation, the parameters were very small for saving the running time. We set $p=2$, $q=3$, both were very small numbers. And the number of voxels was 400, it was also small compared to the real data. The simulation was supposed to retain good result if the parameters increased their values and became closer to real data, but further simulation study was needed to be conducted to prove this. And for each parameter setting of simulation, the simulation was needed to run repeatedly several times to check whether the results got close to each other. And the regression model was not applied to the real dataset we had now yet, it needed to be tested with real data.

This was a good experience for me, I learnt a lot from this and I made a lot of mistakes. Though brain imaging analysis is difficult for me to understand at the beginning, I should keep working on this, keep reading papers to get more knowledge for better understanding, keeping learning new methods and testing them. But when I was struggled with something difficult to understand or could not run the program I wrote successfully, I shouldn't give up so early by myself. This attitude made me stop for quite a long time without making any progress. Actually

when I picked what I learned and did again, I found better comfortable about this. I felt myself

understand a lot of issues what I was confused about at first. But there was not much time for

me. I really feel very regretful about this.

Reference

[1] V.D. Calhoun, T.Adali, G.D. Pearlson, and J.J. Pekar. (2001). A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. *Human Brain Mapping, 14,* 140-151.

[2] V.D. Calhoun, T.Adali, L.K. Hansen, J.Larsen, J.J Pekar. (2003). ICA of Functional MRI Data: An Overview. *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003).*

[3] Elena A.Allen, Erik B.Erhardt, Eswar Damaraju, William Gruner, Judith M. Segall, Rogers F.Silva, Martin Havlicek, Srinivas Rachakonda, Jill Fries, Ravi Kalyanam, Andrew M. Michael, Arvind Cprihan, Jessica A. Turner, Tom Eichele, Steven Adelsheim, Angela D.Bryan, Juan Bustillo, Vincent P. Clark, Sarah W.Feldstein Ewing, Francesca Filbey, Corey C.Ford, Kent Hutchison, Rex E.Jung, Kent A.Kiehl, Piyadasa Kodituwakku, Yuko M.Komesu, Andrew R.Mayer, Godfrey D.Pearlson, John P.Philips, Joseph R.Sadek, Michael Stevens, Ursina Tuscher, Robert J.Thoma and Vince D.Calhoun. (2011). A Baseline for the Multivariate Comparison of resting-state networks. *Frontiers in Systems Neuroscience*. doi: 10.3389/fnsys.2011.00002.

[4] V.D Calhoun, J.J Pekar, V.B. McGinty, T.Adali, T.D. Waston, and G.D. Pearlson. (2002). Different Activation Dynamics in Multiple Neural Systems During Simulated Driving. *Human Brain Mapping, 16,* 158-167.

[5] P.Comon. (1994). Independent Component Analysis, A New Concept? *Signal Processing, 36.3*, 287-314.

[6] Aapo Hyvarinen and Erkki Oja. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Networks, 13,* 411-430.

[7] Erik Barry Erhardt, Srinivas Rachakonda, Edward Bedrick, Elena Allen, Tulay Adali and Vince D.Calhoun. (2011). Comparison of Multi-Subject ICA Methods for Analysis of fMRI Data. *Human Brain Mapping, 32(12)*, 2075-2095.

[8] Calhoun V.D, Liu J, Adal T. (2009). A Review of Group ICA for fMRI Data and ICA for Joint Inference of Imaging, Genetic, and ERP data. *NeuroImage, 45(1s1)*, 143-172.

[9] Ying Guo, Giuseppe Pagnoni. (2008) A Unified Framework for Group Independent Component Analysis for Multi-Subject fMRI data. *NeuroImage, 42*, 1078-1093.

[10] Christian F. Beckmann, Marilena DeLuca, Josepn T.Devlin and Stephen M.Smith. (2005). Investigations into Resting-State Connectivity Using Independent Component Analysis. *The Royal Society, 360,* 1001-1013.

[11] Vince D. Calhoun, Kenta A. Kiehl, and Godfrey D. Pearlson. (2008). Modulation of Temporally Coherent Brain Networks Estimated Using ICA at Rest and During Cognitive Tasks. *Human Brain Mapping, 29(7),* 828-838.

[12] D Glahn. (2000). SPM fMRI Analysis.

[13] N.Correa, T.Adali, Yi-Ou Li and V.Calhoun. Comparison of Blind Source Separation Algorithms for fMRI using a new MATLAB toolbox: GIFT. *ICASSP, Vol 5,* 401-404.

[14] E.Allen, E.Erhardt, T.Eichele, A.R. Mayer, and V.D. Calhoun. (2010). Comparison of pre-normalization methods on the accuracy of group ICA results. *Human Brain Mapping*.

[15] Y.Du, Y.Fan. (2013). Group Information Guided ICA for fMRI Data Analysis. *NeuroImage, 69,* 157-197.

[16] Y. Guo. (2011). A general probabilistic for group independent component analysis and its estimation methods. *Biometrics, 67(4),*1532-1542.

[17] Hong B, Pearlson GD, Calhoun VD. Source-density driven independent component analysis approach for fMRI data. *Hum. Brain Map. 2005;25(4),* 297–307.

[18] Bowman, F.B., Guo, Y. and Derado G. Statistical Approaches to Functional Neuroimaging Data. *Neuroimaging Clinics, 17(4)*, 2007.

[19] Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging. 2004;23(2),*137–152.

[20] Xi-Nian Zuo, Clare Kelly, Jonathan S. Adelstein, Donald F. Klein, F. Xavier Castelllanos, Michael P. Milham. Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression method. *NeuroImage, 49,* 2163-2177.

Appendix A:

Covariate	Spatial maps	Spectra
TEI	13	9
CTQ	18	12
BDI	6	10
EDS	13	.
AGE	11	18
PTSD_diagnosis	8	8
PSS	13	8

Table 1

Appendix B:
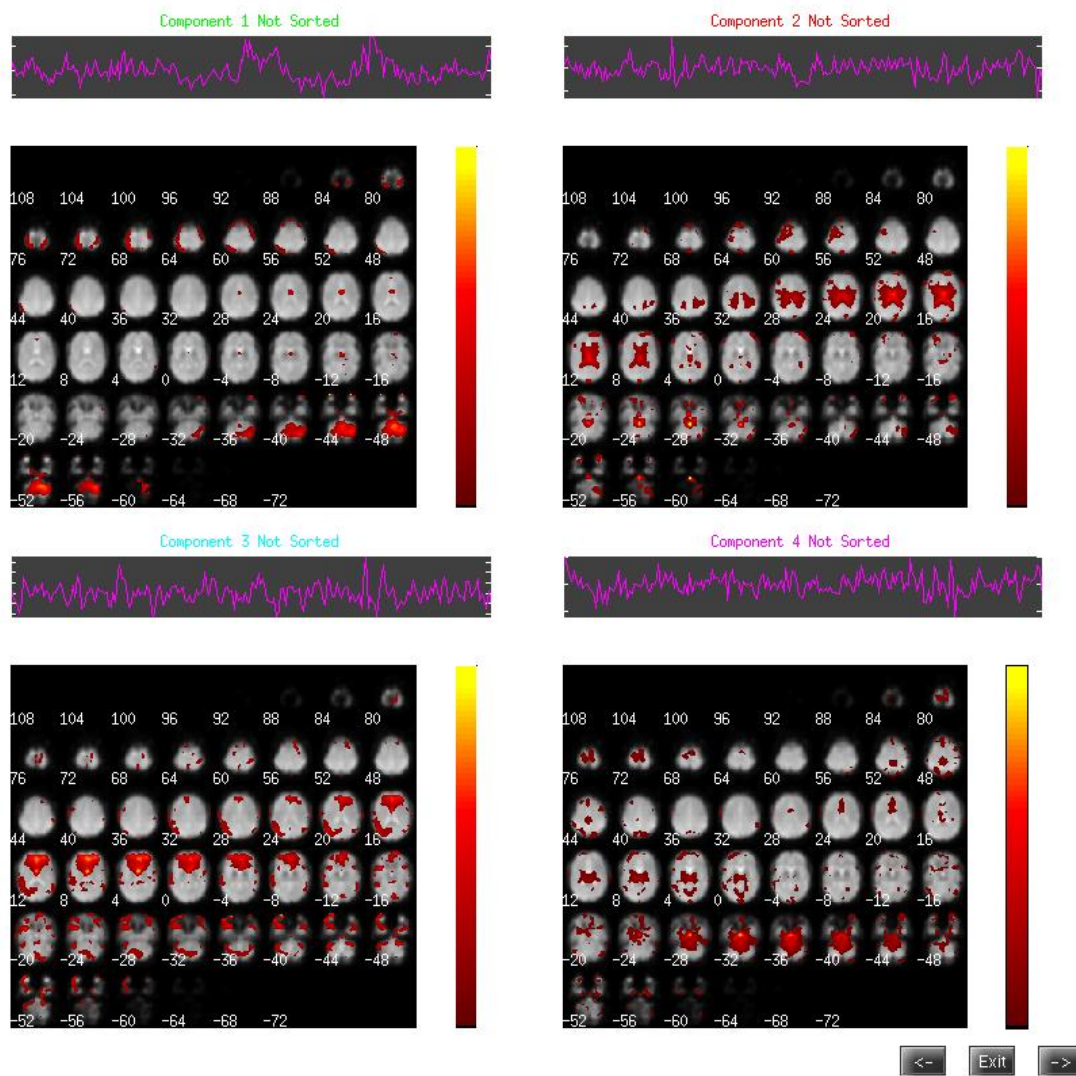
Figure 1
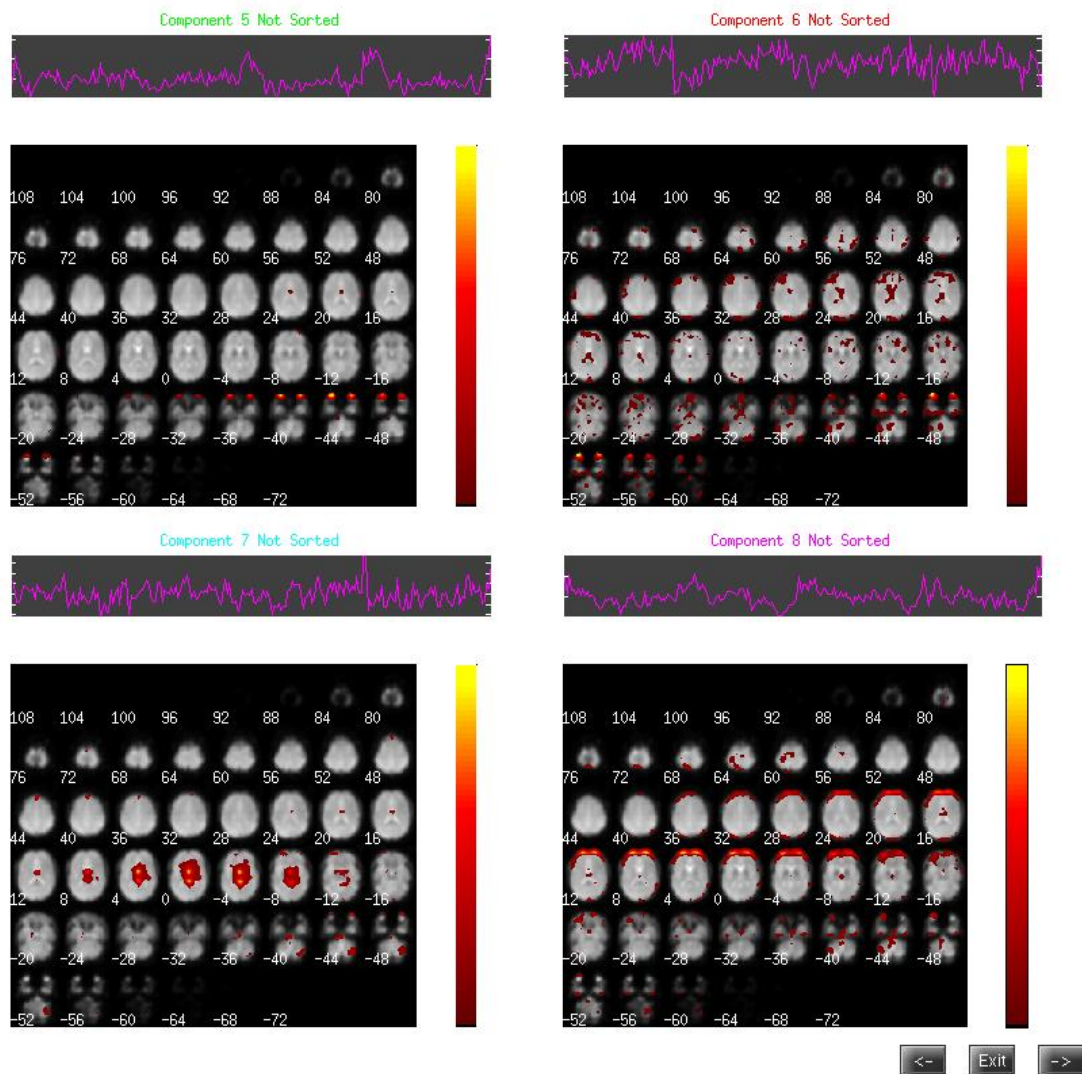
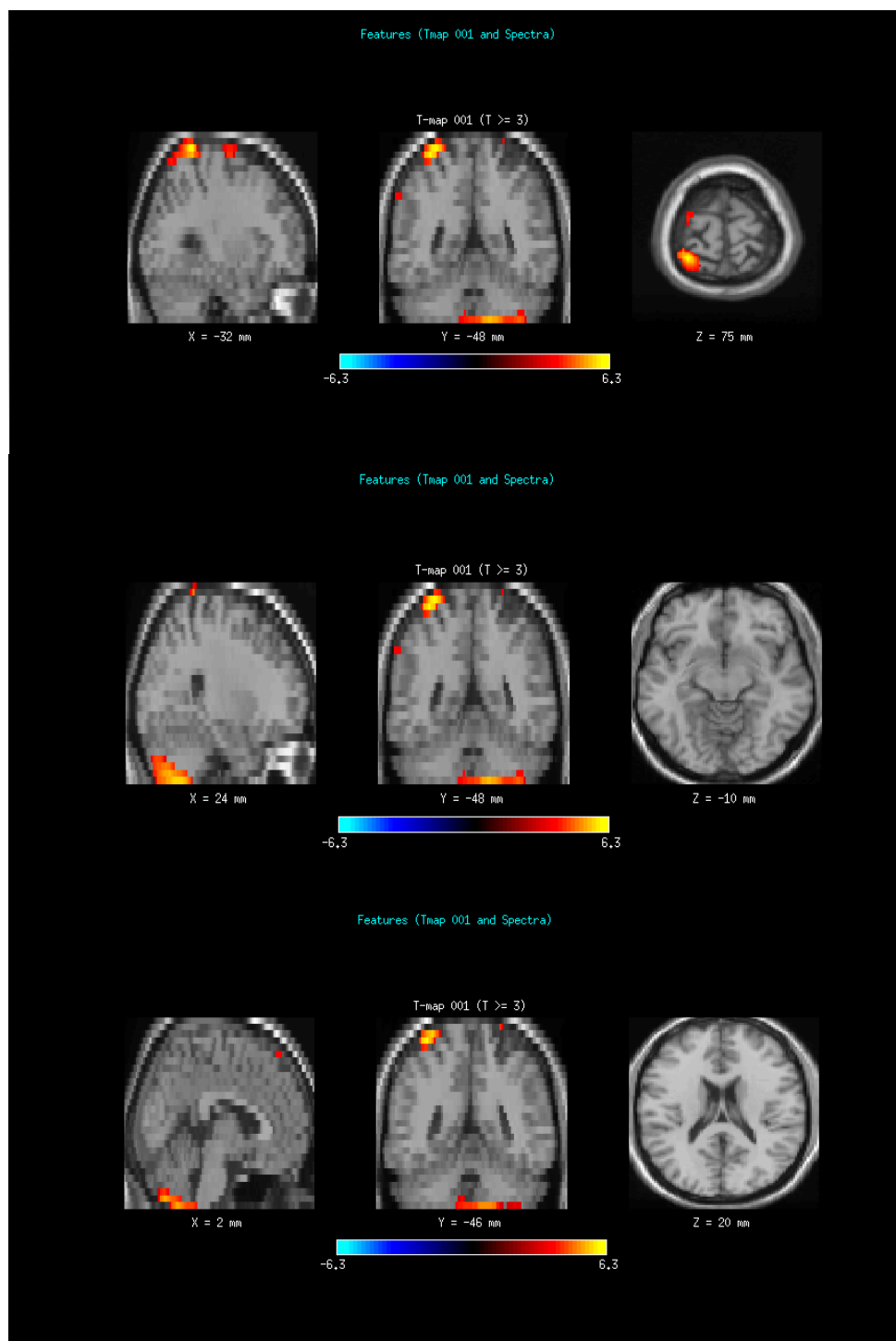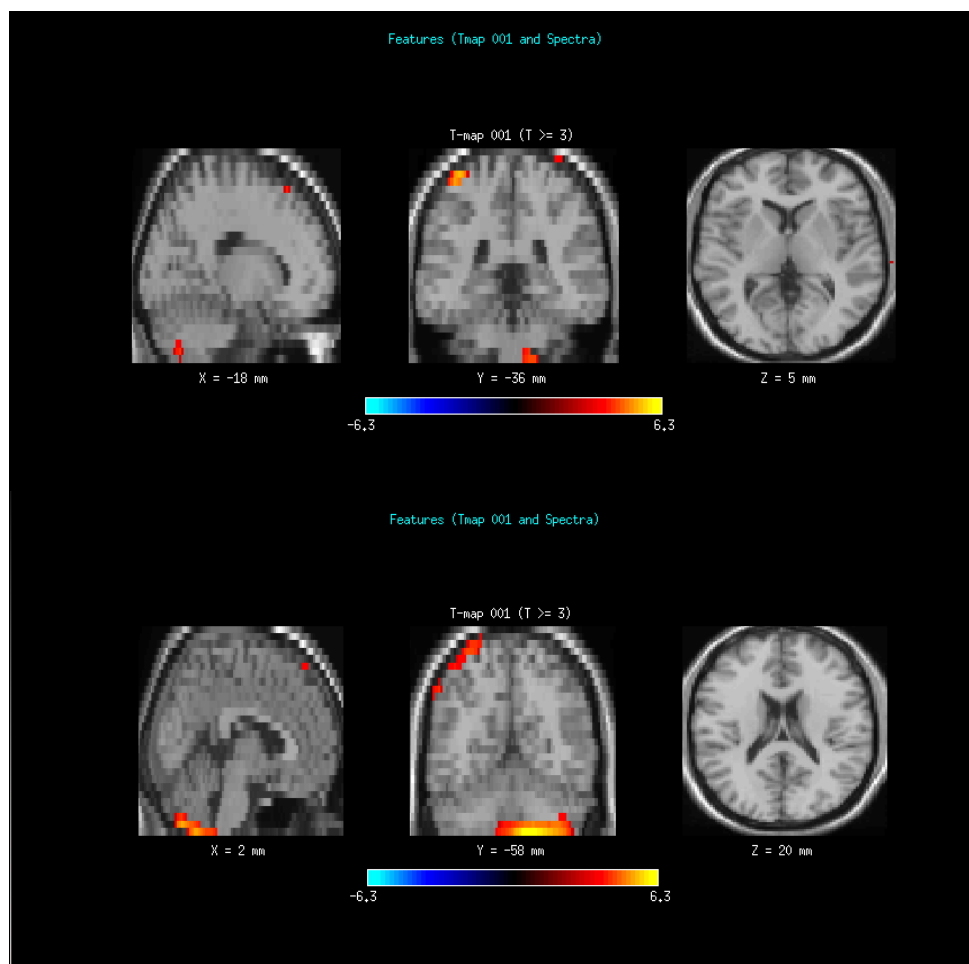**Figure 2**
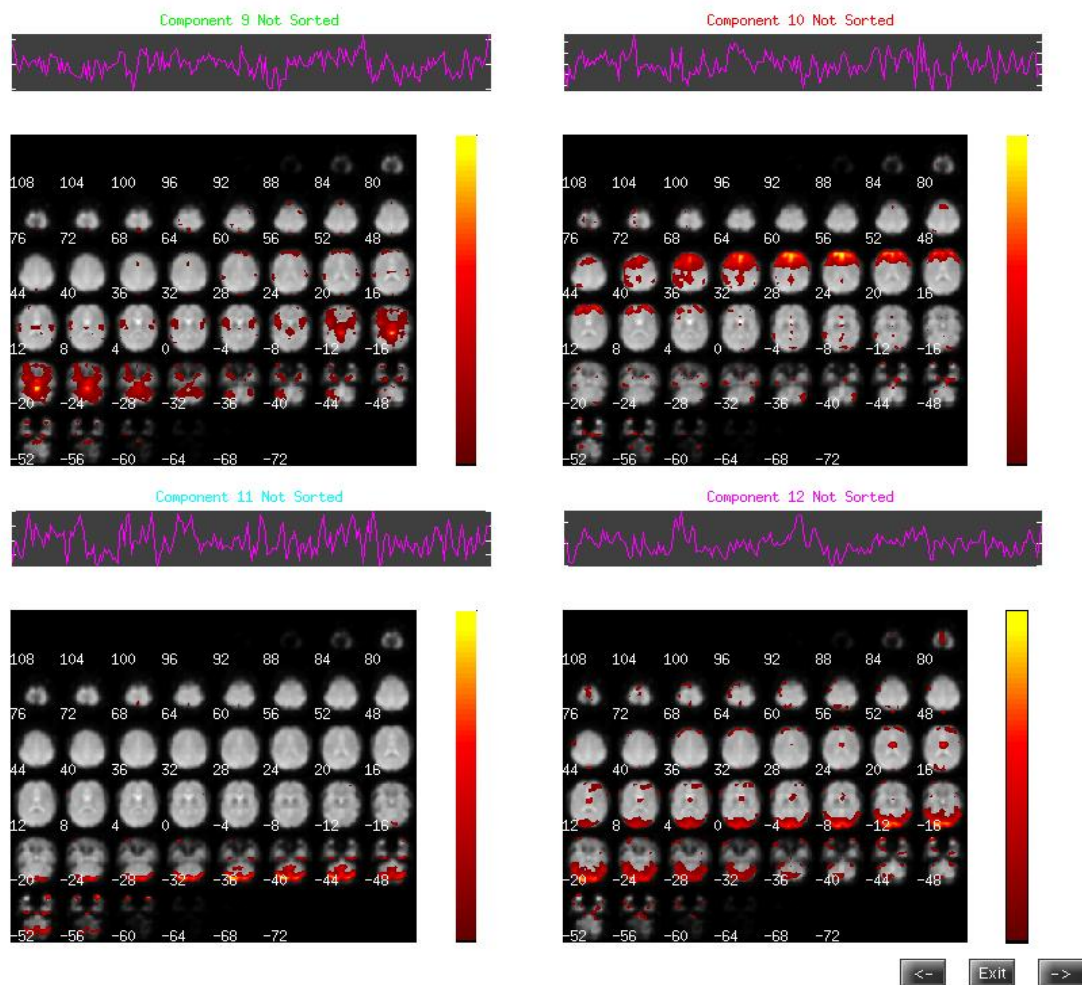
Figure3
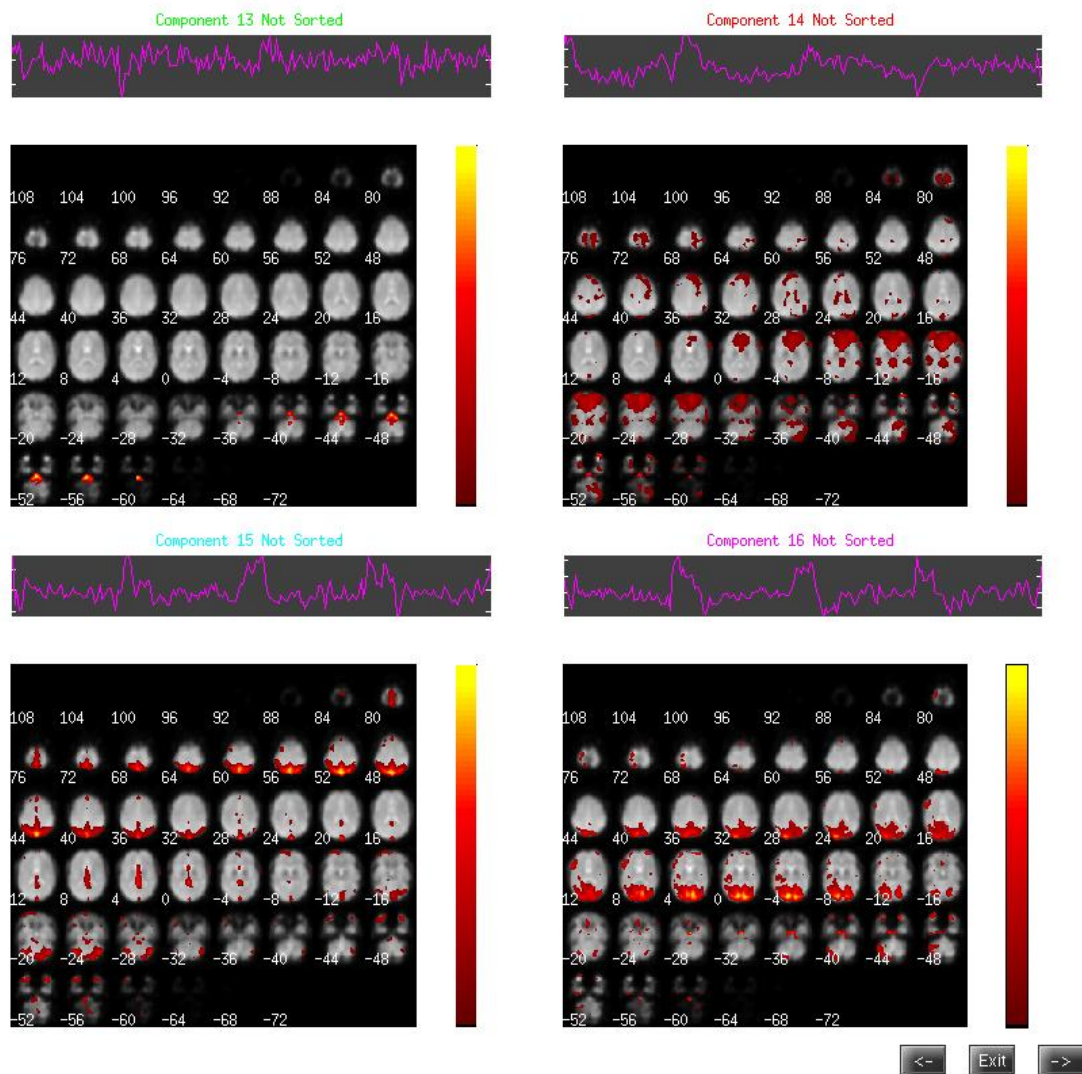
Figure 4
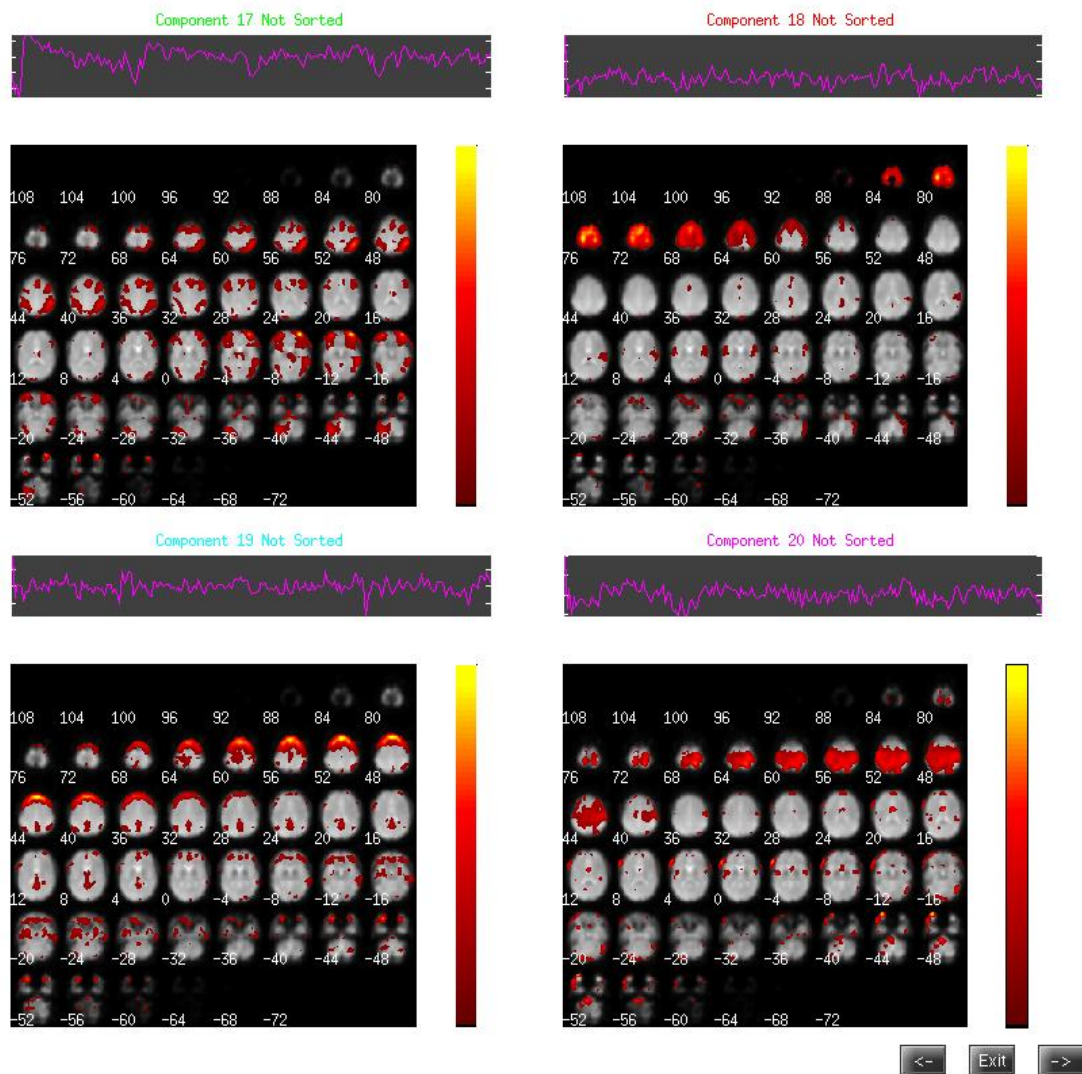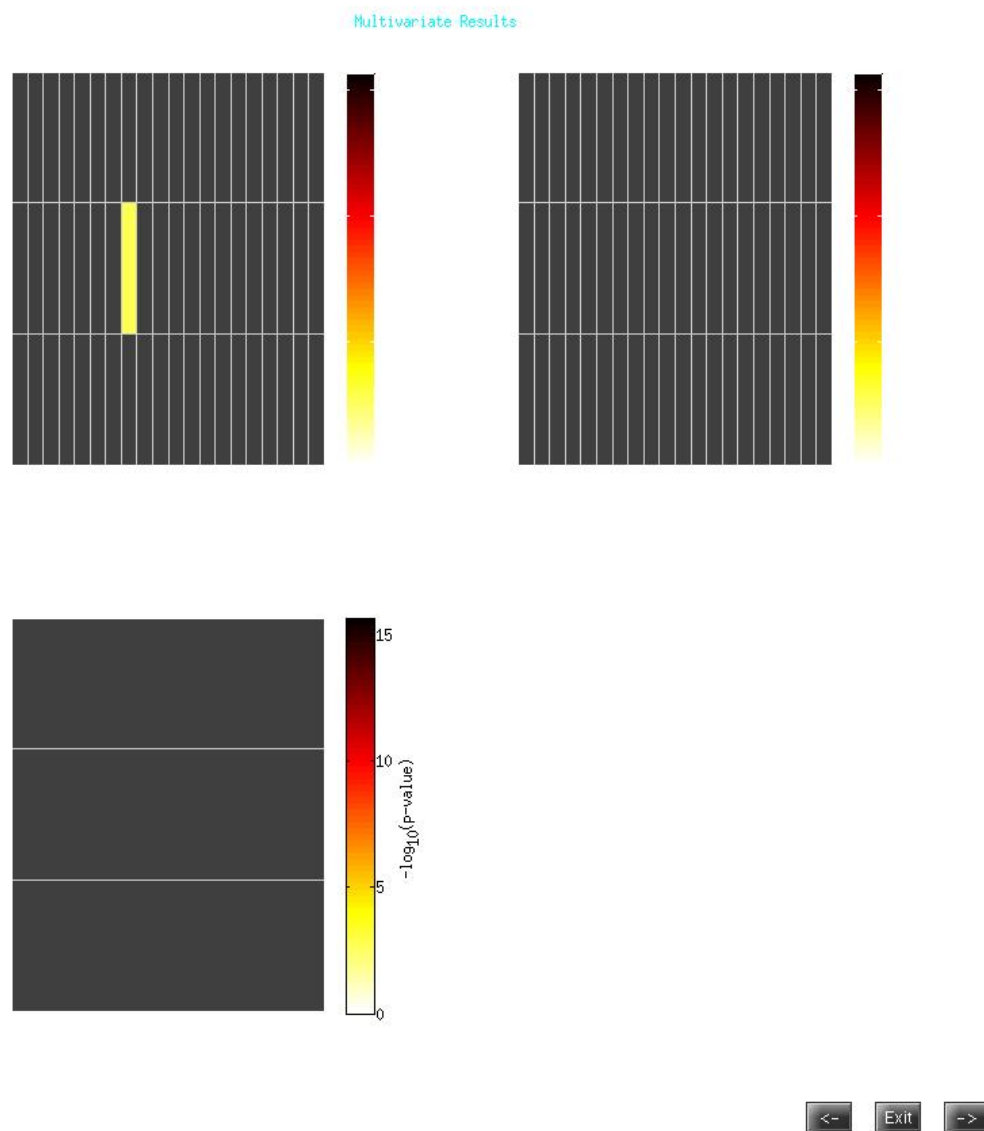
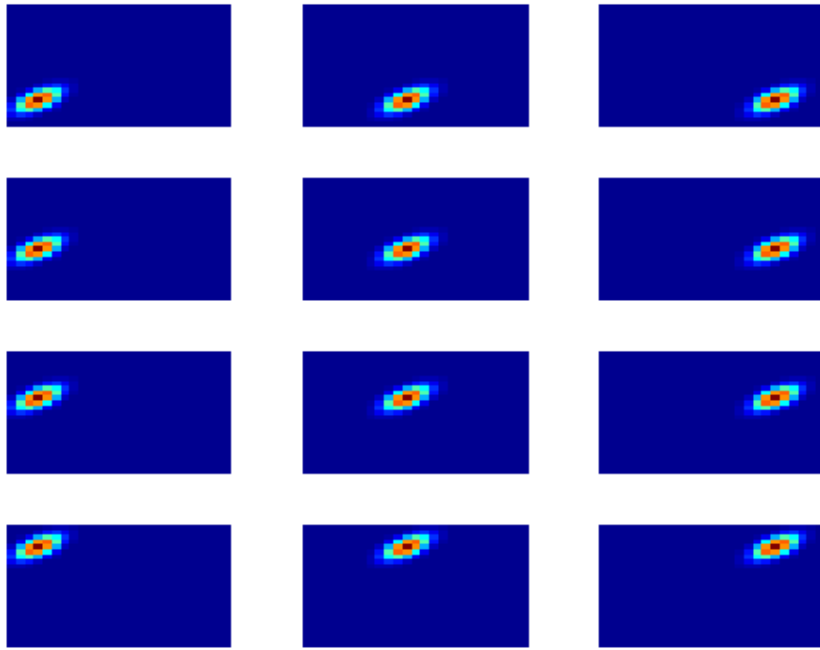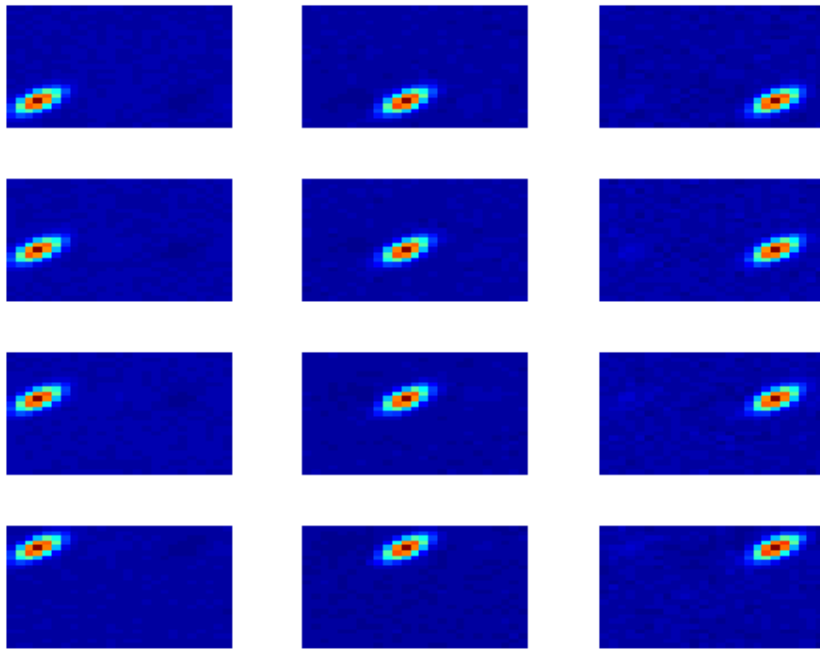**Figure 5**

Figure 6

Figure 7

Figure 8

Figure 9(True covariates)

Figure 10 (Estimated covariates)